(Almost) All Data is Absent Data

Karly Ross* University of Calgary
Pratim Sengupta† University of Calgary.
Wesley Willett‡ University of Calgary

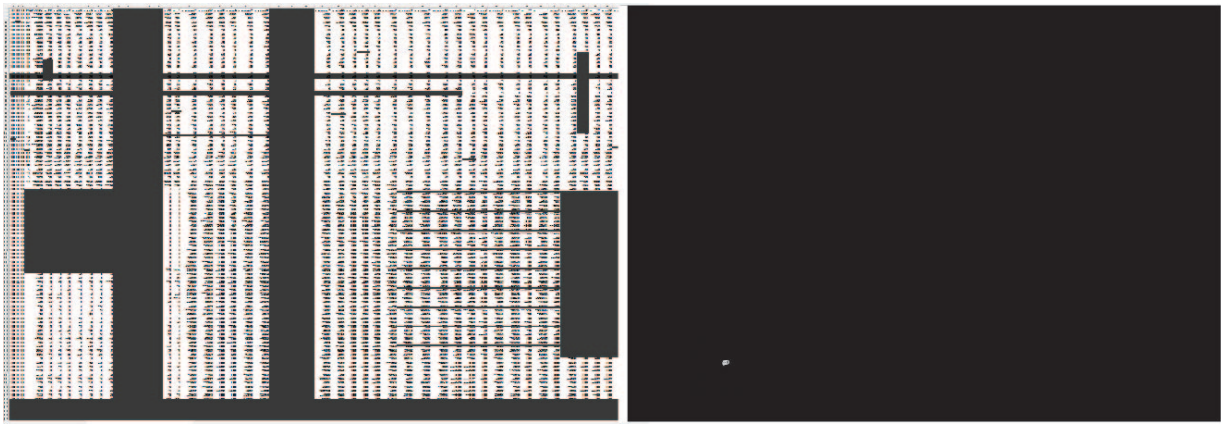

Figure 1: Figure 1 A: This image on the left shows an abstracted data set with the areas with no data blanked out in grey. There are grey columns representing things that have not been collected, rows in grey representing items that have not been counted and various other grey areas and patches with various degrees of patterning. This model has many voids, but all within the existing data structure. Figure 1 B: In this model on the right, a tiny speck of white is in a void. This speck indicates all the data that is collected in what we perceive to be an infinite field of all the data that could be collected on a given subject or topic. This is the *data in a void* model where a small speck of data exists in an infinite void of absent data

Abstract

We explain our model of *data in a void* and contrast it with the idea of *data voids* to explore how the different framings impact our thinking on sustainability. This contrast supports our assertion that how we think about the data that we work with for visualization design impacts the direction of our thinking and our work. To show this we describe how we view the concept of *data in a void* as different from that of *data voids*. Then we provide two examples, one that relates to existing data about bicycle mobility, and one about non-data for local food production. In the discussion, we then untangle and outline how our thinking about data for sustainability is impacted and influenced by the *data in a void* model.

Index Terms: Data, Visualization, Sustainability, Community Process, Axiology, Data Absences

1 Introduction

"It matters what matters we use to think other matters with; it matters what stories we tell to tell other stories with; it matters what knots knot knots, what thoughts think thoughts, what descriptions describe descriptions, what ties tie ties. It matters what stories make worlds, what worlds make stories."

— Donna J. Haraway, Staying with the Trouble: Making Kin in the Chthulucene [7]

What do we think when we think about data and how does this impact our creation of visualization tools for community-level climate change mitigation and adaptation? We approach this question by addressing the idea of data voids and contrast it with our framing of *data within a void*. We use this comparison to discuss ongoing work related to sustainability.

One common framing supporting data production today is that needed data is largely available and that this availability is peppered with *data voids*. We call this the *data voids* model where the voids are usually presented as the small spaces where data has yet to exist (Figure 1 A). In contrast, in our *data in a void* model (Figure 1 B) we suggest that infinitesimal amounts of data exist in an infinite void and that using this as a thinking tool helps us frame the work we choose to do. When we assume data is mostly absent, it reminds us to make deliberate choices about our orientations when creating or using data to build a sustainable and just present and future.

2 Related Works

The idea of *data voids* spans academic and non-academic policy-focused literature. The term *data voids* comes directly from work by Golebiewski and boyde who use it to describe the potential security risks of missing data in internet searches [1]. Another notable example can be found in the Data to Policy Navigator which is a tool for data and policy development [12]. A key suggestion from the navigator is that sometimes there are "cracks" or "gaps" in data and that the types of data gaps can be clustered into four broad groupings. The Data to Policy Navigator positions absent data as something that exists in cracks or gaps as something that can be filled in and made complete if we can simply identify what we are missing, and so for us it appears to be situated in thinking we associate with the *data voids* model.

We adopt the perspective that data may arise due to epistemic and axiological orientation. Simply put, data may not exist because we have yet to imagine them or because we do not find them valuable. In alignment with Correll [4], we adopt the perspective that data is not value- nor politically-neutral. Other work that aligns with this comes from Indigenous scholars such as Walter and Anderson who argue for viewing data construction as rooted in axi-

*karly.ross@ucalgary.ca
†e-mail: pratim.sengupta@ucalgary.ca
‡e-mail: wesley.willett@ucalgary.ca

ology [13], and critical scholars such as Haraway have argued for viewing data through the lenses of standpoint [7]. Furthermore, Drucker has suggested that how we think of data linguistically matters, suggesting that data is what we capture about the world, and thus should be called capta [6]. Missing data and its connection to systems of power is likewise addressed by D'Ignacio and Klein [5]. Elsewhere, ways of noting absence as information are presented in recent work by Sherman and coauthors who outline multiple ways that thinking with data absence can support meaning-making with the deeply-biased data that currently exists [10].

Designing for absence is also a theme that appears in recent work. Song and Szafir address visualization techniques for showing missingness largely aligned with *data voids*, but do not address data that is needed or yet to be imagined [11]. Elsewhere, Andy Kirk highlights that showing small amounts of data in large amounts of space can be a meaningful way to make a point [9]. However, this framing still generally focuses on the gaps and absences in existing data sets, and so is still closer to viewing data as abundant and voidness as the exception. In this way, our line of thinking is strongly influenced by data feminism where our suggestion of thinking about *data in a void* is intended to remind us of power relations in the construction of data.

Opportunities for working with imagined but not yet existing data or data that functions across epistemologies are plentiful. Bressa and coauthors provide a framework for thinking about input visualizations designed to take user input that may make new data [2]. Such methods provide opportunities for us as researchers to think about tools for representing data that have yet to be created. Additionally, we see work spanning epistemic divides in data representation in Kassam and coauthors' international development research on ecological calendars — visual that tools are used to bridge Indigenous knowledge and time scales with scientific knowledge [8]. Informative here is that ecological calendars use existing data that starts from vastly different places in a void which is then connected to support life in changing systems. This further aligns with Walter and Anderson who ask quantitative researchers to consider the values that ground their data construction and what influences those values. They also suggest that rather than looking to Indigenous methodologies as something to reproduce, white-stream and other non-Indigenous researchers may become richer and deeper thinkers by considering the things that often go unnamed when constructing their quantitative data [13].

3 The Model of *Data in a Void*.

The *data voids* model (Figure 1 A) suggests an abundant field of data on most things we could be curious about, with a few voids scattered here and there. In this model, voids have many names such as voids [1], gaps, or cracks [12] or absences [9]. We offer a different model to think about data we do not have, one that does **not** rest on the assumption that **not** having data is an oddity.

Our *data in a void* model recognizes conceptually that even the biggest datasets capture only narrow and discrete slices of the phenomena they purport to measure. They are limited by many things such as the spatial resolution of sensors, the human effort required to observe and document a phenomenon, and the underlying limitations of the technical and social systems that generate them. Furthermore, they (almost) always capture that data in relative isolation, absent the multitudes of possible other complementary data that might help to contextualize and more deeply understand them. Thus (almost) all data is absent data.

Additionally, the *data in a void* model requires us to consider which direction we want to travel. When we think of data specks existing in an infinite void, it is possible to notice that there may be other specks of *data in a void* which we may not fully understand when we are standing squarely in our data. This then requires us to act with deliberate attention if, and when, we want to understand how data exists in other places in a void. So, we suggest that thinking of data as infinitesimal specs in an infinite void helps us to maintain a pluralistic view of data; it requires deep attentional work to consider what, where, and how we want to create new data, or to visualize the data we already have. Therefore, our view of data as existing in a void is axiologically oriented as it impacts how we may consider visualization design for sustainability.

	A	B	C	D	E	F	G	H
1	NAME	ADDRESS	WEB	longitude	latitude	location	Indoor Stalls	Outdoor Stalls
2	James Short Parkade	112 5 Ave SW	https://ww	-114.064	51.0486	(51.04859	44	0
3	City Centre Parkade	221 9 Ave SW	https://ww	-114.068	51.0445	(51.04449	104	0
4	McDougall Parkade	720 5 Ave SW	https://ww	-114.077	51.04907	(51.04906	16	0
5	Lot 72	615 3 Ave SW	https://ww	-114.074	51.05042	(51.05042	12	0
6	Centennial Parkade	608 9 Ave SW	https://ww	-114.075	51.04517	(51.04516	27	22
7	Convention Centre Par	727 1 St SE	https://ww	-114.061	51.04597	(51.04596	8	0

Figure 2: This is a screenshot of the complete "Public Bike Parking" data set in Excel. This data set is from The City of Calgary's open data portal [3].

4 Case Studies

We provide two case studies to think about *data in a void*. The first example is based on an existing data set about bicycle mobility. In our second example, we provide a thought experiment based on the first author's recent experience with local food production and climate adaptation.

4.1 Case Study 1: Bike Parking in the City of Calgary

The City of Calgary, like many other cities, has a data set that represents bike parking [3]. This dataset (Figure 2) documents 211 indoor and 22 outdoor bike parking stalls in downtown parkades — a tiny fraction of the total bike parking in the city. This data fails to represent the many thousands of publicly- and privately-owned bike racks in the city. There is also the effectively infinite space of informal bike parking infrastructure which includes anywhere that a bicycle user can park and/or lock a bike on a rack or elsewhere. More importantly, the data does not capture any information related to the quality, usability, comfort, safety, accessibility, or availability of this parking or its sustainability implications.

This leaves us with a data set that produces more questions than answers. For example, we look at this data and ask: what is the distribution of secure bike parking around the city for purposes beyond workplace commute? Are there areas of use such as around libraries, transit hubs, and daily services like grocery stores that are particularly well or poorly served by such infrastructure? We can also ask questions like what conditions encourage people to use a bike and feel safe parking it. The point here is not that bike parkades are not important; the point here is that knowing where to park a bike downtown in a parkade is a far more niche use case than these other questions. If we only look for gaps in the existing data set, we may only look to add points that are similar in kind — other bike parkades. If we expand the definition to bike parking (rather than bike parkades), we may have far broader coverage, which at the same time, may also be more equitable. The distinction between these two perspectives also has a bearing on sustainability, as adopting the more expansive approach can also help us understand why and how cycling and parking behaviours may change from person to person and across hyper-local contexts.

4.2 Case Study Two: Gardening as a Perpetual Novice

Gardening at individual and community scales is often touted as a means of increasing food security and as a strategy for adaptation to climate change. However, the first author, Ross, suggests that the complexities of growing food in a changing climate are not minor footnotes but essential elements of understanding how our plant-human systems interact. Ross has two and a half decades of food

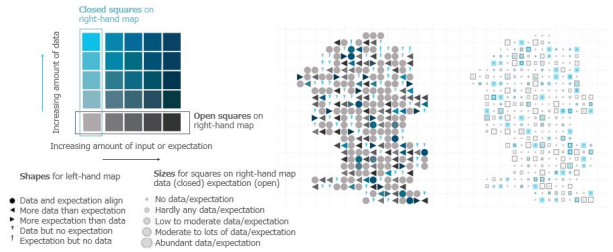

Figure 3: Bivariate glyphs (left) and aligned marks (right) highlight expectation differences in a tile grid map of Calgary communities. The data for these examples is randomly generated to emulate the full set of combinations that could occur using the sample lexicons.The goal of this exploration of designs is to be able to show areas where there are absences across both community data and administrative data.

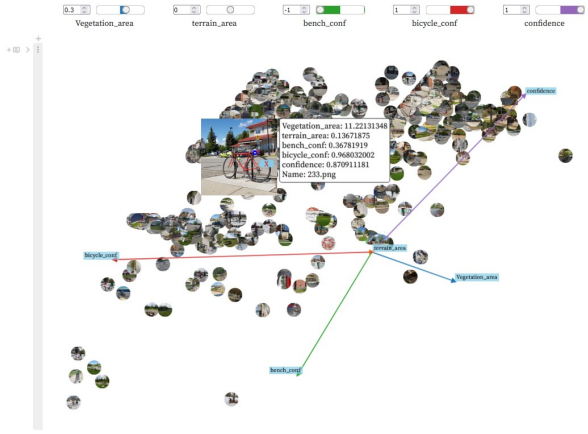

Figure 4: A screenshot of the in-progress implementation of a photo-vis tool that we will use to ground conversations to address infrastructure **and** data absences. In this case, the photo-data used is counter data we are generating to support community knowledge-building and meaning-making about bicycle parking [5]. At this time our generation of counter data is more connected to *data voids*. However, the ongoing design process is intended to provide users tools to communicate about both data as we conceive of it, and data that they know may know we need, but that we have not identified as valuable or thought about. To encourage this we have included visualization functionality where users can change the positions of the photographs based on classification in five categories of relevance for bike parking. They can also add annotations.

production experience in both gardens and on farms and the privilege of growing food in the same space as three family generations. A common view of sustainable food production is that experience and practice lead to becoming a skilled producer. However, as the climate changes the seasonal variability increasingly makes every year new. Climate change produces a sense of being a novice of the local landscape every year.

This spring, in Ross's garden, multiple kale plants grew from the stems of last year's plants. This is a phenomenon never seen in this yard before. There are two questions this brings up, first, how do we know? The knowing that this is rare comes from the systems of family sharing and knowledge transfer which exists outside the processes of academic knowledge and data production. The second question is how do we adapt our food production and knowledge to climate change? None of this information exists in data, or as data as far as we know, it is micro-level information that is essential to the ongoing use of a small piece of land for food production.

5 Discussion

The *data in a void* model provides a tool for thinking about how the existence and absence of data are produced by the values of social/technological systems. It is relevant in many domains including, but not limited to, sustainability. For example, we are currently playing with two approaches for working with present and absent data in relation to bicycle infrastructure. We are also starting to generate ideas about data for local food production.

Our most extensive work to this point has been focused on bicycle infrastructure. In one project we are creating a visual lexicon that connects and compares people's knowledge about a place with existing data. One round of ideation that come from this work is tile mapping which shows interactions between community knowledge and administrative data found in Figure 3. Additionally, we are developing processes to create photographic counter-data [5] and visualizations about bicycle parking. Through this we hope to make space for a pluralistic understanding of data and data absence. To support this goal one step we are taking is the ongoing development of a visualization tool that uses photographs and foregrounds user tinkering and annotation (Figure 4). So, in each case, we first locate ourselves in the data we can access or create, and then we begin to create visualization processes that take us out of that existing data towards either seeing absence more clearly, or to imagining new directions for data construction. Like Sherman and coauthors, noticing absence is a core feature in our meaning-making with data [10].

The second case study points to a connection between sustainable food production and *data in a void*. In this case, the desired and needed data does not exist. However, weather, climate, and generalized gardening data do. This demonstrates that we may have lessons to learn from projects like ecological calendars [8] which combine different epistemic viewpoints in visual data representation. We may also have lessons to learn about our plural ways of knowing by paying attention to the land. While the visualization tools that may be useful are still not fully formed in our minds, there are possible directions which we can imagine. For example, input visualization tools that build on, and also make visual sense of, traditional gardeners' journals, or tools that help us notice calendar dates as well as the seasonality and sequencing of garden events which in turn help us make sense of plant survival over many annual cycles [8]. Both at individual and more regional levels such processes **may** help local food producers to work together to adapt to increased seasonal variability while maintaining sustainable and local food supplies.

Our thinking on sustainability can be influenced by the *data in a void* model which requires us to emphasize that data completeness is impossible, we can only ask if the data we are dreaming of using — or creating —- is pointing us in the direction of survival, life, justice, and the world we want to build. It also encourages us to ask if the tools we build are sufficiently adaptive to help us use and learn from them today and in the near future. When we only look to fill gaps in existing data we may miss noticing what absence is telling about the processes needed to make our present more livable and just and our future more sustainable.

Acknowledgments

The authors wish to thank Kath Blair, Desmond Larsen-Rosner, and Abena Yeboah for their help in clarifying our writing. We also wish to thank Yash Lanjewar for his amazing work on the development of the photo-vis tool. We are also grateful to all the members of the Data Experience Lab for their collaboration and input.

REFERENCES

[1] D. Boyd and M. Golebiewski. Data voids: Where missing data can easily be exploited, 2018. 1, 2

[2] N. Bressa, J. Louis, W. Willett, and S. Huron. Input visualization: Collecting and modifying data with visual representations. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–18, 2024. 2

[3] City of Calgary. Public bicycle parking. https://data.calgary.ca/Transportation-Transit/Public-Bicycle-Parking/fbvs-aj5g/about_data. 2

[4] M. Correll. Ethical dimensions of visualization research. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–13, 2019. 1

[5] C. D'ignazio and L. F. Klein. *Data feminism*. MIT press, 2020. 2, 3

[6] J. Drucker. Humanities approaches to graphical display. *Digital Humanities Quarterly*, 5(1):1–21, 2011. 2

[7] D. J. Haraway. *Staying with the trouble: Making kin in the Chthulucene*. Duke University Press, 2020. 1, 2

[8] K.-A. S. Kassam, M. L. Ruelle, C. Samimi, A. Trabucco, and J. Xu. Anticipating climatic variability: The potential of ecological calendars. *Human Ecology*, 46:249–257, 2018. 2, 3

[9] A. Kirk. The design of nothing: Null, zero, blank. In *OpenVis Conference talk, YouTube*, vol. 28, 2014. 2

[10] J. Sherman, R. Morrison, L. Klein, and D. Rosner. The power of absence: Thinking with archival theory in algorithmic design. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, pp. 214–223, 2024. 2, 3

[11] H. Song and D. A. Szafir. Where's my data? evaluating visualizations with missing data. *IEEE transactions on visualization and computer graphics*, 25(1):914–924, 2018. 2

[12] United Nations Development Program. Identify data gaps, 2023. 1, 2

[13] M. Walter and C. Andersen. *Indigenous statistics: A quantitative research methodology*. Taylor & Francis, 2013. 2